

# Natural Language Processing in the Indian Context

Mehul Agarwal

October 2018

**Abstract.** Natural Language Processing (NLP) is a field under the umbrella of computer science and more specifically Artificial Intelligence. It deals with a computer's understanding of complex human languages. It includes speech recognition, understanding and generation. In recent times, NLP has become popular in applications such as Voice Assistants, Chatbots and Customer Support. It has three main components- input, processing and output. On input side it deals with reading data from texts, images and understanding human handwriting. This has to be followed by the computationally hard problem of processing data as human interaction is often imprecise, grammatically incorrect, loaded with slangs and acronyms and varies with context. In the end, it generates a natural language as output. Additional challenges appear in input and output when dealing with speech over text. Growth of NLP has been aided by advances in Artificial Intelligence (especially Machine Learning and Neural Networks), Computational Linguistics and increase in computer processing speeds and storage space.

Almost all of language applications are in English and this limits their scope of use in multilingual countries like India, where most citizens primarily speak native languages. This paper explores development of NLP applications for eighteen common Indian languages with a special focus on eight of the most spoken languages- Bengali, Gujarati, Hindi, Malayalam, Punjabi, Tamil, Telugu and Urdu. Each of these languages have their own linguistic obstacles due to their structure and semantics. The paper attempts to provide pragmatic and viable use cases for NLP in India and addresses challenges and shortcomings for the same. Among the use cases discussed are: translating Hinglish (a blend of Hindi and English), automating replies to Right To Information (RTI) queries which manually take thirty days on average and transcribing documents in regional courts.

The ideas mentioned in the paper bring the fruits of technology to all citizens and help us bridge the digital divide in the country.